

Using 1 trillion files helps scientist find a needle in a haystack

June 22, 2018

[Using 1 trillion files helps scientist find a needle in a haystack](#)

by Bradley Wade Settlemyer

A Los Alamos National Laboratory scientist having trouble solving a stubborn research problem needed some help – his scientific simulations had generated a sea of data, but it took so long to search the data that he couldn't find the information he needed. He found himself looking for the proverbial needle in a haystack. At the same time, the lab's storage research team had been hard at work on another classic big data problem: creating massive numbers of files as quickly as possible. The day the team met with the scientist, you could say that Big Science and Big Data put their heads together – and now they're making history.

The Ultrascale Systems Research Center in the lab's High Performance Computing Division is tasked with realizing the next generation of supercomputing. With efforts in storage research, novel computer architectures and extreme scale platform management, the center is uniquely positioned to tackle these seemingly impossible computing challenges. In particular, a collaboration between the center and Carnegie Mellon University had developed an experimental file system designed to support unprecedented numbers of files and folders. It wasn't obvious it would work, but it seemed like a chance worth taking.

This story first appeared in [Albuquerque Journal](#).

Managed by Triad National Security, LLC for the U.S Department of Energy's NNSA